Journal of Clinical Epidemiology 69 (2016) 5–7

Journal of Clinical Epidemiology

EDITORIAL

A mega - issue

This month's edition is larger than usual so that we can catch up with a backlog of accepted manuscripts. We view this as a positive sign of receiving increasing numbers of good manuscripts.

Guidelines are discussed in three articles. McGowan et al. report on the experience of implementing a world bank loan to develop a clinical practice guideline program in Kazakhstan; this was quite complex, given the paucity of resources in Russian. It is heartening that health systems foreign aid loans from institutions such as the world bank to lower/middle income countries now include building systems for evidence – based decision making.

When developing guidelines, how worried should we be about accepting the effect size from a single well designed RCT – should the confidence in this be rated as low? Gartlehner et al. challenge the results of widely cited papers (some in this journal), on the 'proteus phenomenon', the greater tendency in science for early replications of a work to contradict the original findings. Gartlehner found this was not true in a random sample of 100 Cochrane reviews that included at least three RCTs across a range of clinical topics; the exception was where there was a large effect size in the first trial.

Developing guidelines for reporting case series need more attention as case series are still clinically important for a number of situations. Examples are delayed adverse effects and health technologies that get modified too rapidly for trials to keep up with them. Bing et al. propose a 20 item checklist; this resulted from a Delphi of experts assessing 105 case series analysed by principal component analysis followed by an expert panel.

Prediction Rules are addressed in four articles.

A commentary by Steyerberg and Harrell propose a new combination of internal and external validation steps for the validation of prediction models that makes the most of limited sample sizes.

Adoption and clinical use of robustly validated evidence- based clinical decision rules, continues to be a challenge. Logistics, complexity and lack of transparency are the most frequent reasons cited for reluctance of clinicians to use them. Sanders et al. report on the trade-off between simplifying and the poorer resulting discrimination of simplifying a cardiac decision rule the Emergency Department Assessment of Chest pain Score (EDACS). It is accepted that there needs to be external validation of diagnostic and prognostic indexes before they should be used in clinical practice; Riley et al. show how multivariate meta-analysis of individual participant data helped the calibration of prediction models in cancer and deep vein thrombosis.

Giovanni studied the added value of multiple testing. They tackled the problem of how to predict length of stay, cognitive function and discharge destination [home or institution] in a cohort of over 300 patients admitted to 10 geriatric centres on whom a battery of tests was conducted. The clinicians' clinical prediction was accurate for the extremes of functioning but was poor for those in the middle – and the battery of tests did not add materially to this accuracy.

Questionnaires appear to be increasingly applicable. For example, evidence is accumulating that Patient Reported Outcomes can be obtained from children as young as six; Denbaek et al. show over 80% agreement between children aged 6-14 and parents on reporting illness-related absenteeism from school.

However, parsimony – shortening, to minimise the respondent burden on all patients – is critical. But how to do it? Two papers in this issue address this. Guillemin suggest the following criteria: (1) Document the validity of the original scale; (2) Take the conceptual model into account; (3) Preserve content validity; (4) Preserve psychometric properties; (5) Document the justification for selection of each item; and (6) Validate the short-form CMS in an independent sample. These authors provide a good example of the challenges in meeting these criteria as applied to the challenge of shortening (from 43 to 20 items) a composite outcome Patient Reported Outcome Scale the Mini-OAKHQOL a disease specific outcome instrument for knee and hip osteoarthritis; one key aspect, demonstration of the responsiveness, is still needed.

Quality of life scales need to be easily interpretable if they are to be useful for providing estimates of the magnitude of benefit and harm for practice or policy. Two of the most widely used QOL Instruments that have been used in thousands of trials are the SF36 and European Organisation for Research and Treatment of Cancer [EORTC] QLQ-30 questionnaires-these have 8 and 15 outcomes respectively and it is recommended that these be reported individually. The SF36 does have 2 summary component scores Physical Component Summary Mental Component Score but these do not perform well psychometrically and the FDA no longer accepts them as primary/major outcomes in pivotal

http://dx.doi.org/10.1016/j.jclinepi.2015.12.001

studies for labelling purposes. In this issue Aaronson and colleagues utilise best-practice methods that emphasize the goodness of fit with the different aspects of factor analysis dimensionality laid out in 8 higher order models. They conclude that a summary score for the EORTC QLQ-C30 is robust and can be used.

Statistical and related methods issues are the focus of five articles. Briel et al. report on a large follow up study of eight Research Ethics Boards in Switzerland, Germany and Canada to assess the reasons for early stopping of trials. They found that most discontinuations of clinical trials were not based on pre-planned interim analyses or stopping rules; they recommend these be included in all protocols and reporting guidelines such as CONSORT. Mavridis et al. using a database of trials in schizophrenia, argue that network meta-analysis can be used to adjust for both a] publication bias and b] small-study effects − two major threats to the validity of meta-analysis.

The area under the curve statistic is the standard for quantifying the ability of a risk model to discriminate between individuals who will or will not manifest the outcome of interest. A risk model with a higher AUC will be able to better separate the predicted risk distribution curves of events and non-events. AUC lacks an important parameter: the incidence of the outcome in the population, and thus, it is not well suited for clinical decision. Campbell et al. solve this with use of a new statistic, the prediction impact curve (PIC)-this estimates the percentage of events prevented when a risk model is used to assign high-risk individuals to an intervention; they demonstrate this with an example. of the prevention of coronary heart disease.

Competing risk of dying [i.e. the patient dies but not of the condition of interest; they die instead from another unrelated disease] has been overlooked in many studies in leading journals; Walraven shows that a third of 100 Kaplan Meier Risk Estimates in 2013 were biased upwards due to this. They recommend the use of the cumulative incidence function to overcome this.

Kollhorst present nice example using previous prescribing practices to demonstrate the challenge in choosing an instrument variable to handle confounding by indication.

The World Health Organisation measure of drug exposure Defined Daily Dose is widely used in pharmacoepidemiology studies; Sinnott et al. show that it will result in underestimation and overestimation of association with harms and advocate the ‘days supply’ as being more accurate.

Research wastage is becoming a recurring theme in JCE. The evidence of ongoing ‘research wastage’ continues to accumulate. Extra attention to this needed. Sawin et al. carried out an in 2011 update of their 2004 study that showed that 70-75% of trials fail to be cited in subsequent trials and that this is more often the negative trials that are missing. This is despite the updated CONSORT statement recommendation that new results be set within the context of the existing evidence. Biased and inadequate citation of prior research in reports of cardiovascular trials is a continuing source of waste in research.

Observational studies are the focus of four studies: Collecting data in longitudinal cohort studies is expensive and there is little guidance published on methods of optimal ‘cost-efficiency’ regarding when and how often to collect data. Using data from the East-West study, the Finnish part of the Seven Countries Study of cardiovascular epidemiology, Reinikainen et al. describe the data and modelling used to show that most items only need collecting every 10 years.

As precision medicine combines the study of both the genotype and phenotype, Mendelian randomisation needs to be better understood to appreciate the increasingly common problem of type 1 errors in genetics studies. As Burgess et al. state ‘‘There is a distinction between Mendelian randomization as it was initially conceived and performed (mainly for circulating biomarkers using few genetic variants in relevant gene regions) and how it is often used today (often opportunistically using large numbers of genetic variants whose functional relevance is unknown)’’.. In order to filter out the large numbers of false positive statistically significant [often highly so] associations they propose using an adaptation of the classic Doll and Hill criteria for causation- that will surely be reassuring to clinical epidemiologists as a nice change from the jargon in genetic studies.

Powels et al. show that the quality of reporting of confounding in leading epidemiology and general medicine journals, has improved only marginally 3-5 years after the publication of the STROBE guideline. The authors suggest that authors be required to submit a completed checklist and be expected to carry out quantitative bias analysis for unmeasured confounding.

RCTs and their variants are the focus of four articles. A new term ‘In silico’ is introduced for the first time to a JCE title in this issue − it means ‘performed on computer or via computer simulation’; Wikipedia tells us the phrase was coined in 1989 as an allusion to the Latin phrases in vivo, in vitro, and in situ, which are commonly used in biology (see also systems biology) and refer to experiments done in living organisms, outside of living organisms, and where they are found in nature, respectively. Using the example of sumatriptin trials for migraine headaches, Chabaud et al. report on the use of a RCTs model with 100,000 Monte Carlo repetitions to generate a treatment effect in a virtual population of patients obtained after modeling human behavior, disease progress, and drug effects using specific mathematical models and numerical methods. Virtual patients are then randomized in virtual trials considering different designs to allow comparisons of estimated power, accuracy of the estimation of treatment effect, and number of patients receiving active treatment. They compared (A) Parallel, (B) crossover, (C) randomized withdrawal, (D) early escape, (E) play the winner, (F) drop

the loser, and (G) N of 1. We look forward to hearing readers views on this.

Estellat et al. conducted an innovative ethics exercise to ask a panel of independent rheumatologists whether patient profiles derived from real patients selected from the control group of trials of biologics in rheumatoid arthritis were being denied available alternative efficacious therapy. 70% of the patients in the background therapy plus placebo control groups were deemed to be receiving substandard care but these respondents would be prepared to enter half of these same patients into a trial. This raises important aspects of the trade-off between the demands for placebo – controlled trials by approval agencies versus the ethical equipoise in enrolling such patients.

Sample size calculations for stepped wedge and cluster randomised trials are complex and often wrong. Using a new formula, Hemming and Taljaard provide a simpler way of calculating these with useful tables.

Clarke et al. showed that the RCT design can be used to assess the reliability of peer review. Such studies are important since peer review, like democracy, is not ideal but still seems the fairest way we have to make tough choices in science. So it is reassuring to see the high reliability of grant funding decisions in fellowships for early career researchers in their randomised trial. As systematic reviews are increasingly used as representing the state of knowledge, assessing and avoiding risk of bias in reviews deserves high priority.

Whiting et al. present a new generic risk-of-bias tool ROBIS systematic reviews of aetiology, diagnosis, interventions and prognosis. Several tools exist for undertaking critical appraisal and quality assessment of systematic reviews but none specifically aim to assess the risk of bias in systematic reviews; the ROBIS tool was designed to fill this gap in risk of bias assessment. This now needs to be evaluated for its usability and usefulness by the different audiences it is aimed at, namely guideline developers, authors of overviews of systematic reviews (‘‘reviews of reviews’’), and review authors who might want to assess or avoid risk of bias in their reviews.

Not all systematic reviews are equal! Seehra et al. reviewed over 300 systematic reviews published in the Core Clinical Journals in 2014. They found that although some assessment of risk of bias was carried out in over 70%, rarely was a decision made to include only the better designed studies – which casts serious doubt on accuracy and usefulness of the effect size, however carefully computed.

Many teachers of clinical epidemiology still favour NNTB and NNTH [Number needed to treat for benefit/harm] as a good way of communicating the effect size of interventions and this remains a core element in CAML [Critical Appraisal of the Medical Literature]. In addition to the influence of prevalence, the reader needs to pay attention to the accuracy of estimates of the outcome [CAML]; van Werkhoven et al. provide a nice example of what happens when outcomes are missed, and the substantive effect upon the NNTB /NNTH when identifying Community Acquired Pneumonia episodes in a population-based pneumococcal vaccination trial.

Presentation of results of complex interventions is not easy – Hutchings et al. using the exemplar of two gastrointestinal complex interventions [modernising endoscopy services using 10 different qualitative and quantitative research methods; an RCT comparing the clinical and cost-effectiveness of doctors and nurses undertaking upper and lower GI endoscopy in 23 endoscopy units], present a nice clean and clear way of presenting the different conclusions to the different stakeholders – with separate summary tables for patients, providers, and health system managers. It would be interesting to extend this to providing separate Summary of Findings Tables in Cochrane Reviews for different stakeholders.

Reporting and methodological quality of 197 published surgical meta-analyses found in Medline in 2013 was found to be quite variable with few satisfying g the PRISMA and AMSTAR criteria; it was reassuring to see the correlation between the scores on the two scales achieved a R^2 of 0.79 suggesting that the same attributes are being assessed by each.

Many organisations have spent large amounts on moving to a paper e free environment but the clinical and laboratory systems do not harmonise. Lim Faye et al. provide an example of pediatric infections in a hospital that demonstrate the potentially harmful pitfalls of incomplete data integration where 30% of microbiological tests never made it onto the patients health record. Validation should be required when introducing of any such system both for clinical care and research.

Peter Tugwell
J. André Knottnerus
E-mail address: Laura.Tugwell@uottawa.ca (P. Tugwell)